Robust Tests in Regression Models With Omnibus Alternatives and Bounded Influence

Lan WANG and Annie QU

A robust approach for testing the parametric form of a regression function versus an omnibus alternative is introduced. This generalizes existing robust methods for testing subhypotheses in a regression model. The new test is motivated by developments in modern smoothing-based testing procedures and can be viewed as a robustification of a smoothing-based conditional moment test. It is asymptotically normal under both the null hypothesis and local alternatives. The robustified test retains the "omnibus" property of the corresponding smoothing test; that is, it is consistent for any fixed smooth alternative in an infinite-dimensional space. It is shown that the bias of the asymptotic level under shrinking local contamination is bounded only if the second-order Hampel's influence function is bounded. The test's performance is demonstrated through both Monte Carlo simulations and application to an agricultural dataset.

KEY WORDS: Bounded influence; Conditional moment test; Influence function; Local contamination; Omnibus alternative; Regression; Robust test; Smoothing.

1. INTRODUCTION

Testing the validity of a specified model is important for regression analysis. A long-standing concern for model specification tests is their sensitivity to unusual observations in the data, which often can have destructive effects. For instance, in simple linear regression, the classical F test statistic can be distorted by a single outlier. Therefore, developing model lack-of-fit tests that are robust to outlier contamination is crucial.

Considerable effort has been devoted to the development of robust testing procedures. For linear regression models, Schrader and Hettmansperger (1980) introduced the ρ_c test based on Huber's M estimators; Sen (1982) proposed an aligned M test; Markatou and Hettmansperger (1990) introduced an aligned generalized M test; He, Simpson, and Portnoy (1990) discussed power and level breakdown properties of certain tests; and Markatou and He (1994) proposed testing procedures based on one-step high–breakdown point bounded influence estimators. More recently, Heritier and Ronchetti (1994), Markatou and Manos (1996), Agostinelli and Markatou (2001), and Liu, Markatou, and Tsai (2005) developed robust tests for general parametric models and nonlinear regression models.

We refer to the aforementioned procedures as classical robust tests because they are restricted to testing nested hypotheses. In other words, they require the specification of a "larger" alternative model. In practice, such a "larger" alternative model is often chosen for mathematical convenience. The classical tests are designed to be powerful for this particular alternative but may fail to uncover important model violations if even the "larger" alternative model does not adequately fit the data.

Testing procedures with the so-called "omnibus" property have received much attention recently (see Hart 1997 for a general review). The omnibus tests are attractive because they are consistent against a large class of alternatives that satisfy certain smoothness conditions. These include, in particular, the class of smoothing-based *conditional moment tests* of Härdle and Mammen (1993), González-Manteiga and Cao (1993), Zheng (1996), Li and Wang (1998), Fan, Zhang, and Zhang (2001), and Horowitz and Spokoiny (2001), among others. A common feature of all of these tests is that their test statistics can be expressed as smoothing-based nonparametric estimators of certain population moments, which are zero under the null hypothesis and strictly positive under any fixed alternative. The smoothing-based tests do not need to assume a parametric distribution for the random errors and thus are robust against the mispecification of the error distribution. However, they are not robust against outliers, because a linear local average of the response variable is not robust against outliers (Härdle 1992, chap. 6). The sensitivity of ordinary smoothing-based tests to outlier contamination is demonstrated in Section 5.

We propose an effective way to robustify a smoothing-based conditional moment test. Our test bases the inference on the centered *asymptotic rank transformation* (defined in Sec. 2) of the residuals from a robust fit under the null hypothesis. A representative example of the robustified Zheng test (Zheng 1996) is thoroughly discussed in the subsequent sections. We prove that the proposed test preserves the omnibus property of the regular smoothing test as in the Zheng test. Thus it is especially appropriate for applications in which one needs to diagnose model adequacy but there is not sufficient a priori information for specifying the form of the model under the alternative.

We study local robustness through influence function analysis. The classical tests can be (asymptotically) written as $T(F_n)$, where F_n is the empirical distribution function of the data. For smoothing-based tests, they are most appropriately expressed as bivariate "plug-in" functionals $T(F_h, F_n)$, where F_h is a smoothing-based estimator of the underlying distribution. A direct result of this new feature for smoothing-based tests is that their second-order von Mises (1947) functional expansion consists of terms of different converging rates, because F_h converges to the underlying distribution function more slowly than F_n does. It is shown that the robustified test has first-order influence of zero and second-order influence function bounded in the response direction. The bias of the asymptotic level under shrinking local contamination is bounded only if its second-order Hampel (1974) influence function is bounded.

The rest of the article is organized as follows. Section 2 introduces the test statistics. Section 3 discusses large-sample properties of the robust test under the null hypothesis, fixed alternative, and local alternatives, and Section 4 analyzes the robust-

Lan Wang is Assistant Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: *lan@stat.umn.edu*). Annie Qu is Associate Professor, Statistics Department, Oregon State University, Corvallis, OR 97331 (E-mail: *qu@stat.orst.edu*). The authors thank two anonymous referees, the associate editor, and the joint editor, whose insightful comments significantly improved the manuscript. The research of Qu was supported in part by the National Science Foundation (grant DMS-03-48764).

Journal of the American Statistical Association
March 2007, Vol. 102, No. 477, Theory and Methods
DOI 10.1198/016214506000001130

ness property through influence function analysis. Section 5 reports numerical simulations and an agricultural dataset analysis, and Section 6 provides generalizations and a discussion. The Appendix provides technical proofs.

2. PRELIMINARIES AND NOTATION

Assume that we observe independent and identically distributed random vectors $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$, where $\mathbf{X}_i$ is a $l \times 1$ vector and Y_i is a scalar. We consider regression model $Y_i = r(\mathbf{X}_i) + \epsilon_i$, where $r(\cdot)$ is an unknown smooth function and the ϵ_i's are iid random variables with mean 0 and variance σ^2, that are independent of the $\mathbf{X}_i$'s.

A question of fundamental importance in regression is whether the conditional mean $E(Y_i|\mathbf{X}_i) = r(\mathbf{X}_i)$ can be modeled parametrically. Although modern smoothing methods offer an alternative approach for estimating $m(\cdot)$, parametric models remain very popular among practitioners, because they provide a succinct summary of the underlying physical phenomena and can be estimated efficiently if they are specified correctly. To evaluate the lack-of-fit of a proposed model, the following hypotheses are of interest:

$$H_0 : r(\mathbf{x}) \in S_\Theta \qquad \text{versus} \qquad H_1 : r(\mathbf{x}) \notin S_\Theta, \tag{1}$$

where $S_\Theta = \{m(\cdot, \theta), \theta \in \Theta\}$ is a parametric family of functions, Θ is a subset of a Euclidean space, and $m(\cdot, \theta)$ is a function on $\mathcal{R}$.

In the sequel, we assume that the random vector $(\mathbf{X}_i, Y_i)$ has joint distribution function $F(\mathbf{x}, y)$ and joint density function $f(\mathbf{x}, y)$, $\mathbf{X}_i$ has marginal distribution function $F_{\mathbf{X}}(\mathbf{x})$ and marginal density function $f_{\mathbf{X}}(\mathbf{x})$, and Y_i given $\mathbf{X}_i$ has conditional density function $f_{Y|\mathbf{X}}(y|\mathbf{x})$. The random error ϵ_i has distribution function $H(\cdot)$ and density function $h(\cdot)$. Under H_0, there exists some $\theta_0 \in \Theta$ such that $P(r(\mathbf{X}_i) = m(\mathbf{X}_i, \theta_0)) = 1$.

2.1 The Zheng Smoothing Conditional Moment Test

The idea of moment tests can be traced back to Newey (1985) and Tauchen (1985), who discussed a very general framework for testing the specification of maximum likelihood models. In regression settings, the current state of the art is to estimate the population conditional moments using nonparametric smoothing methods.

A representative example of smoothing-based conditional moment tests for diagnosing the mispecification of a parametric regression has been given by Zheng (1996). To test (1), Zheng proposed using an estimator of the population moment $E[\epsilon_i(E(\epsilon_i|\mathbf{X}_i)f_{\mathbf{X}}(\mathbf{X}_i))]$ [or, equivalently, $E(E^2(\epsilon_i|\mathbf{X}_i)f_{\mathbf{X}}(\mathbf{X}_i))$] as a test statistic, because this moment condition is zero under the null hypothesis and strictly positive under any alternative. More specifically, the density function $f_{\mathbf{X}}(\mathbf{X}_i)$ is estimated using the kernel method

$$\widehat{f}_{\mathbf{X}}(\mathbf{X}_i) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^l} K\left(\frac{\mathbf{X}_i - \mathbf{X}_j}{h}\right). \tag{2}$$

Similarly, the conditional expectation $E(\epsilon_i|\mathbf{X}_i)$ is estimated by

$$\widehat{E}(\epsilon_i|\mathbf{X}_i) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^l} K\left(\frac{\mathbf{X}_i - \mathbf{X}_j}{h}\right) \epsilon_j \big/ \widehat{f}_{\mathbf{X}}(\mathbf{X}_i). \tag{3}$$

The outer-layer expectation in the moment condition is then estimated by a sample average. The estimators in (2) and (3) are often called "leave-one-out" kernel estimators because the ith observation is left out. (See Härdle 1992 for more discussion of kernel-based nonparametric estimation methods.) Finally, ϵ_i is estimated by $\widehat{\epsilon}_i = Y_i - m(\mathbf{X}_i, \widehat{\theta}_n)$, which are residuals under the null hypothesis, and $\widehat{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ_0. Assembling those pieces together, we obtain the Zheng test statistic

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^l} K\left(\frac{\mathbf{X}_i - \mathbf{X}_j}{h}\right) \widehat{\epsilon}_i \widehat{\epsilon}_j. \tag{4}$$

Zheng established that when $h \to 0$, $nh^l \to \infty$, and H_0 holds,

$$nh^{l/2} V_n \to \mathrm{N}(0, \tau^2), \tag{5}$$

where $\tau^2 = 2\sigma^4 \int K^2(u)\, du \int f_{\mathbf{X}}^2(\mathbf{x})\, d\mathbf{x}$. The test statistic $nh^{l/2} V_n$ converges to infinity for any fixed alternative, and the test can detect local alternatives converging to the null at the rate $n^{-1/2} h^{-l/4}$.

2.2 Asymptotic Rank Transformation

The central idea of robustifying the smoothing-based conditional moment test is to use rank-transformed residuals. Given $\widehat{\epsilon}_1, \ldots, \widehat{\epsilon}_n$ residuals from a fit under the null hypothesis, we call $n^{-1} \sum_{i=1}^{n} I(\widehat{\epsilon}_i \leq \widehat{\epsilon}_j)$ the *asymptotic rank transform* of $\widehat{\epsilon}_j$, where $I(\cdot)$ is the indicator function. Note that $\sum_{i=1}^{n} I(\widehat{\epsilon}_i \leq \widehat{\epsilon}_j)$ gives the rank of $\widehat{\epsilon}_j$ among all of the n residuals. We propose applying Zheng's test to the centered asymptotic rank transform of residuals, $\widehat{\epsilon}_j^* = \frac{1}{n} \sum_{i=1}^{n} I(\widehat{\epsilon}_i \leq \widehat{\epsilon}_j) - \frac{n+1}{2n}$, $j = 1, \ldots, n$. The robustified version of V_n is

$$V_n^* = \frac{1}{n(n-1)} \sum_{j_1=1}^{n} \sum_{j_2 \neq j_1} \frac{1}{h^l} K\left(\frac{\mathbf{X}_{j_1} - \mathbf{X}_{j_2}}{h}\right) \widehat{\epsilon}_{j_1}^* \widehat{\epsilon}_{j_2}^*.$$

The new test is similar in spirit to that of Bickel (1978), but the Bickel test is consistent only toward nested alternatives, whereas the robustified conditional moment test will be shown to retain the property of being consistent against each member of a very large class of alternatives. The asymptotic normality discussed in the next section requires only that the regression parameter be $\sqrt{n}$-consistently estimated under H_0. The influence analysis in Section 4 further reveals that for the foregoing test to have the desired robustness property, the residuals must come from a robust fit.

3. ASYMPTOTIC DISTRIBUTIONS

The large-sample properties of V_n^* are studied through its asymptotic distributions under the null hypothesis, fixed alternative, and appropriate local alternative sequences. We state the following assumptions:

Assumption 1. The density function $f_{\mathbf{X}}(\mathbf{x})$ of $\mathbf{X}_i$ and its first-order derivatives are uniformly bounded.

Assumption 2. The errors ϵ_i are iid random variables with bounded probability density function.

Assumption 3. The parameter space Θ is a compact and convex subset of a Euclidean space. There exists an estimator $\widehat{\theta}_n$ such that under the null hypothesis, $\sqrt{n}(\widehat{\theta}_n - \theta) = O_p(1)$, whereas under the local alternative sequences given in (7), $\sqrt{n}(\widehat{\theta}_n - \theta^*) = O_p(1)$, where θ and θ^* are both interior points in Θ, a compact and convex set.

Assumption 4. There exists a positive continuous function $M(\mathbf{x})$ such that $\forall\, \theta_1, \theta_2$, $|m(\mathbf{x}, \theta_1) - m(\mathbf{x}, \theta_2)| \le M(\mathbf{x})|\theta_1 - \theta_2|$.

Assumption 5. $K(u)$ is a symmetric density function with compact support and a continuous derivative such that $\int K(u)\,du = 1$. The smoothing parameter h satisfies $h \to 0$ and $nh^l \to \infty$.

These assumptions are essentially the same as those required by Zheng (1996) and are standard in the literature on smoothing-based tests. The smoothness and moments conditions are regular for ensuring the consistency and normality of the parameter estimator under the null model. The assumptions on the kernel function and the smoothing parameter are no more than what are normally required in kernel smoothing estimation.

Under H_0, the test statistic V_n^* is asymptotically equivalent to

$$V_n^{**} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \ne i} \frac{1}{h^l} K\left(\frac{\mathbf{X}_i - \mathbf{X}_j}{h}\right) \times [H(\epsilon_i) - 1/2][H(\epsilon_j) - 1/2]. \quad (6)$$

Because $H(\cdot)$ is the distribution function of ϵ_i, $H(\epsilon_i)$, $i = 1, \dots, n$, have an iid uniform distribution on $(0, 1)$. The result of Zheng's test can be readily applied, and we obtain the following result.

Theorem 1. Given Assumptions 1–5, under H_0,

$$nh^{l/2} V_n^* \to \mathrm{N}\left(0, \frac{1}{72} \int K^2(u)\,du \int f_{\mathbf{X}}^2(\mathbf{x})\,d\mathbf{x}\right).$$

The proof of this theorem is given in the Appendix. An alternative way to derive asymptotic normality under the null hypothesis through influence function analysis is discussed in the next section.

Remark. To use this result to construct an asymptotic test, we need a consistent estimator of the asymptotic variance. We may estimate $\int K^2(u)\,du \int f_{\mathbf{X}}^2(\mathbf{x})\,d\mathbf{x}$ straightforwardly by $[n(n-1)h^l]^{-1} \sum_{i=1}^{n} \sum_{j \ne i} K^2(\frac{\mathbf{x}_i - \mathbf{x}_j}{h})$.

The next theorem presents the asymptotic behavior of V_n^* under any fixed alternative.

Theorem 2. Given Assumptions 1–5, under any fixed alternative,

$$V_n^* \xrightarrow{p} E\big[E^2\big(H\big(\epsilon_i + (r(\mathbf{X}_i) - m(\mathbf{X}_i, \theta_0)) - (r(\mathbf{X}_j) - m(\mathbf{X}_j, \theta_0))\big) - H(\epsilon_i)\big) f_{\mathbf{X}}(\mathbf{X}_i) | \mathbf{X}_i\big],$$

where $\theta_0 = \arg\min_{\theta \in \Theta} E[Y_i - m(\mathbf{X}_i, \theta)]^2$ and $\mathbf{X}_i$ and $\mathbf{X}_j$ are independent copies of random variables from the covariate distribution $f_{\mathbf{X}}(\mathbf{x})$.

The statistic V_n^* converges to zero only if there exists a set with measure 1 such that for any $\mathbf{x}$ in this set, we have

$$E\big[H\big(\epsilon_i + (r(\mathbf{x}) - m(\mathbf{x}, \theta_0)) - (r(\mathbf{x}_j) - m(\mathbf{X}_j, \theta_0))\big) - H(\epsilon_i)\big] = 0.$$

This occurs only when $P(r(\mathbf{X}_j) - m(\mathbf{X}_j, \theta_0) = C) = 1$ for some constant C. It is reasonable to assume that the null class of models is sufficiently general to contain all location shifts in the y direction; in other words, if $m(\mathbf{x}, \theta_0)$ belongs to the null class of models, then so does $m(\mathbf{x}, \theta_0) + C$. Alternatively, we may enlarge the null class of models by including such location shifts. As a result, for any fixed alternative, $nh^{l/2} V_n^* \to \infty$ as $n \to \infty$. Theorem 2 establishes the omnibus property of the test; it will reject any alternative in the infinite-dimensional alternative space when the sample size $n \to \infty$. The asymptotic rank transformation thus preserves the omnibus property of the smoothing-based tests.

Similar to the original smoothing test, our robustified test can detect local alternatives converging to the null at the rate $n^{-1/2}h^{-l/4}$. More specifically, the following local alternative sequence is often under consideration:

$$H_{1n}: m(\mathbf{x}) = f(\mathbf{x}, \theta_0) + n^{-1/2} h^{-l/4} q(\mathbf{x}), \quad (7)$$

where $q(\mathbf{x})$ is continuously differentiable. As the sample size grows larger, this local alternative becomes closer to the null model. Therefore, the idea is to make it harder to reject the null hypothesis when the sample size increases.

Theorem 3. Given Assumptions 1–5, under H_{1n},

$$nh^{l/2} V_n^* \to \mathrm{N}\left(\gamma, \frac{1}{72} \int K^2(u)\,du \int f_{\mathbf{X}}^2(\mathbf{x})\,d\mathbf{x}\right),$$

where $\gamma = [Eh(\epsilon_i)]^2 E[(q(\mathbf{X}_i) - Eq(\mathbf{X}_i))^2 f_{\mathbf{X}}(\mathbf{X}_i)]$.

This rate is slower than $n^{-1/2}$, which usually can be detected by a parametric test. This is the price that the test must pay to be omnibus. However, as h converges to 0 at appropriate speed, this rate can be arbitrarily close to $n^{-1/2}$.

The large-sample results by no means give any indication as to how the test will behave in a finite sample. Its performance in finite sample size is studied through numerical simulation in Section 5.

4. ROBUSTNESS PROPERTIES

4.1 von Mises Functional Expansion

In this section the von Mises analysis (see Fernholz 1983 for a good review) is illustrated, because it not only leads to an alternative approach to derive asymptotic normality under H_0, but also provides the basis on which to calculate the Hampel influence function, which allows us to study the tests' local stability (Sec. 4.2). Both the Zheng test and the robustified Zheng test have degenerate first-order influence. Other examples of zero first-order influence have been given by Hampel, Ronchetti, Rousseeuw, and Stahel (1986, p. 349); Heritier and Ronchetti (1994); and Cantoni and Ronchetti (2001), among others.

If we let $\widehat{G}_n^*$ denote the empirical distribution function of $Y_1 - m(\mathbf{X}_1, \widehat{\theta}), \ldots, Y_n - m(\mathbf{X}_n, \widehat{\theta})$, then the robustified Zheng test statistic V_n^* can be asymptotically expressed as

$$\iint \big[\widehat{G}_n^*\big(y - m(\mathbf{x}, \theta(F_n))\big) - 1/2\big] \times \left(\int \big[\widehat{G}_n^*\big(y_1 - m(\mathbf{x}, \theta(F_n))\big) - 1/2\big] \times \widehat{f}_h(\mathbf{x}, y_1)\, dy_1\right) dF_n(\mathbf{x}, y),$$

where F_n is the empirical distribution function of $(\mathbf{X}_i, Y_i)$, $\theta(F_n)$ is an estimator of θ, and $\widehat{f}_h(\mathbf{x}_i, y_i) = [(n-1)h^{2l}]^{-1} \times \sum_{j\neq i}^n K_1(\frac{y_j - y_i}{h}) K_2(\frac{\mathbf{x}_j - \mathbf{x}_i}{h})$ is a smoothing estimator of the joint density function of $(\mathbf{X}, Y)$, where K_1 and K_2 are two kernel functions. Thus, an appropriate functional for the robustified Zheng test is a bivariate one, $T^*(F, F)$,

$$\iint \big[G^*\big(y - m(\mathbf{x}, \theta(F))\big) - 1/2\big] \times \left(\int \big[G^*\big(y_1 - m(\mathbf{x}, \theta(F))\big) - 1/2\big] f(\mathbf{x}, y_1)\, dy_1\right) dF(\mathbf{x}, y),$$

where F is the distribution function of $(\mathbf{X}_i, Y_i)$ and G^* is the distribution function of $Y_i - m(\mathbf{X}_i, \theta(F))$, that is, $G^*(s) = \iint_{v \le m(u, \theta(F)) + s} dF(u, v)$. Note that the definition of G^* depends on F; if F satisfies the null hypothesis, then $G^*(s)$ becomes $H(s)$, the distribution function of the random error ϵ_i. Therefore, the robustified Zheng test can be asymptotically represented as a plug-in estimator, $T^*(\widehat{F}_h, F_n)$.

Let $A(t) = T^*(F + tG_1, F + tG_2)$, where F is a distribution satisfying H_0, and G_1 and G_2 are two other distribution functions that we specify later. We can expand $T^*(F + tG_1, F + tG_2)$ around (F, F) if we assume that $m(\mathbf{x}, \theta)$ is twice continuously differentiable in θ and that $\theta(F + tG_2)$ is twice continuously differentiable in t. Direct computation yields $A(0) = 0$, $\frac{dA(t)}{dt}|_{t=0} = 0$, and

$$\begin{aligned}
\frac{1}{2}\frac{d^2}{dt^2}A(t)\bigg|_{t=0}
&= \int \Big(\dot{G}^*\big(y - m(\mathbf{x}, \theta(F))\big) f(\mathbf{x}, y)\, dy\Big)^2 d\mathbf{x} \\
&\quad - \iint \dot{G}^*\big(y - m(\mathbf{x}, \theta(F))\big) \\
&\quad \times \int \big[H\big(y_1 - m(\mathbf{x}, \theta(F))\big) - 1/2\big] g_1(\mathbf{x}, y_1)\, dy_1\, dF(\mathbf{x}, y) \\
&\quad - \iint \big[H\big(y_1 - m(\mathbf{x}, \theta(F))\big) - 1/2\big] \\
&\quad \times \int \dot{G}^*\big(y - m(\mathbf{x}, \theta(F))\big) f(\mathbf{x}, y_1)\, dy_1\, dG_2(\mathbf{x}, y) \\
&\quad + \iiint \big[H\big(y - m(\mathbf{x}, \theta(F))\big) - 1/2\big] \\
&\quad \times \big[H\big(y_1 - m(\mathbf{x}, \theta(F))\big) - 1/2\big] \\
&\quad \times g_1(\mathbf{x}, y_1)\, dy_1\, dG_2(\mathbf{x}, y),
\end{aligned} \tag{8}$$

where $g_1(\mathbf{x}, y)$ is the density function corresponding to $G_1(\mathbf{x}, y)$, $\dot{G}^*(y - m(\mathbf{x}, \theta(F)))$ denotes $\frac{d}{dt}G_t^*(y - m(\mathbf{x}, \theta(F + tG_2)))|_{t=0}$, and $G_t^*(s)$ represents $G^*(s)$ under contamination, that is, $G_t^*(s) = \iint_{v \le m(u, \theta(F + tG_2)) + s} d(F + tG_2)(u, v)$. When $t = 0$, $G_t^*(s)$ is simply $H(s)$, the distribution function of the random error ϵ_i.

To derive the asymptotic distribution of the robustified Zheng test under H_0, take $t = 1$, $G_1 = \widehat{F}_h - F$ and $G_2 = F_n - F$ and apply the fact that $\int [H(y - m(\mathbf{x}, \theta(F))) - 1/2]\, dF(\mathbf{x}, y) = 0$ under H_0. A careful analysis shows that the leading term in (8) is the last term,

$$\iiint \big[H\big(y - m(\mathbf{x}, \theta(F))\big) - 1/2\big] \times \big[H\big(y_1 - m(\mathbf{x}, \theta(F))\big) - 1/2\big] \widehat{f}_h(\mathbf{x}, y_1)\, dy_1\, dF_n(\mathbf{x}, y),$$

which is asymptotically equivalent to (6), the approximation in Section 3. Then Hall's (1984) result on degenerate U-statistics can be applied to obtain the asymptotic normality for T^*.

The foregoing technique can also be applied to the Zheng test, because it can be asymptotically expressed as a plug-in estimator of the following bivariate functional, $T(F, F) = \iint (y - m(\mathbf{x}, \theta(F)))(\int (y_1 - m(\mathbf{x}, \theta(F))) f(\mathbf{x}, y_1)\, dy_1)\, dF(\mathbf{x}, y)$. Therefore, the von Mises calculus heuristically suggests the null distribution for the Zheng test, which is equivalent to (5).

4.2 Hampel's Second-Order Influence Function

To investigate the stability of the test under infinitesimal local contamination, we examine the Hampel influence function (Hampel 1974). The influence function of $T(F)$ at $(\mathbf{x}, y)$ is simply $IF(\mathbf{x}_0, y_0; T) = \lim_{\epsilon\to 0} \epsilon^{-1}(T(F_\epsilon) - T(F))$, where $F_\epsilon = (1-\epsilon)F + \epsilon\Delta_{(\mathbf{x}_0, y_0)}$, with $\Delta_{(\mathbf{x}_0, y_0)}$ the point mass function at $(\mathbf{x}_0, y_0)$. For the robustified Zheng test and the Zheng test, the first-order influence is zero. Following Hampel's original definition, if we take $G_1 = G_2 = \Delta_{(\mathbf{x}_0, y_0)} - F$ in the second-order expansion (8), then the four terms will converge at the same rate, leading to the following second-order influence function of the robustified Zheng test:

$$\begin{aligned}
IF^*&(\mathbf{x}_0, y_0) \\
&= \int \bigg[\int \dot{G}^*_{\Delta_{(\mathbf{x}_0, y_0)}}\big(y - m(\mathbf{x}, \theta(F))\big) f(\mathbf{x}, y)\, dy\bigg]^2 d\mathbf{x} \\
&\quad + 2\big[H\big(y_0 - m(\mathbf{x}_0, \theta(F))\big) - 1/2\big] \\
&\quad \times \int \dot{G}^*_{\Delta_{(\mathbf{x}_0, y_0)}}\big(y - m(\mathbf{x}_0, \theta(F))\big) f(\mathbf{x}_0, y)\, dy \\
&\quad + \big[H\big(y_0 - m(\mathbf{x}_0, \theta(F))\big) - 1/2\big]^2,
\end{aligned} \tag{9}$$

where $\dot{G}^*_{\Delta_{(\mathbf{x}_0, y_0)}}(y - m(\mathbf{x}, \theta(F)))$ denotes $\frac{d}{dt}G_t^*(y - m(\mathbf{x}, \theta(F_{2t})))|_{t=0, G_2 = \Delta_{(\mathbf{x}_0, y_0)} - F}$. Straightforward calculations yield

$$\begin{aligned}
\dot{G}^*_{\Delta_{(\mathbf{x}_0, y_0)}}&\big(y - m(\mathbf{x}, \theta(F))\big) \\
&= h\big(y - m(\mathbf{x}, \theta(F))\big)\dot{\theta}(\mathbf{x}_0, y_0, F) \\
&\quad \times \left(\int \frac{d}{d\theta} m(v, \theta(F))\, dF_{\mathbf{X}}(v) - \frac{d}{d\theta} m(\mathbf{x}, \theta(F))\right) \\
&\quad + I\big(y_0 \le y + m(\mathbf{x}_0, \theta(F)) - m(\mathbf{x}, \theta(F))\big) \\
&\quad - H\big(y - m(\mathbf{x}, \theta(F))\big),
\end{aligned}$$

where $\frac{d}{d\theta}m(\mathbf{x},\theta(F))$ is the derivative of $m(\mathbf{x},\theta)$ with respect to θ evaluated at $(\mathbf{x},\theta(F))$ and $\dot\theta(F,G_2)=\frac{d\theta(F+tG_2)}{dt}|_{t=0}$. Similarly, the Zheng test has second-order influence function

$$\begin{aligned} IF(\mathbf{x}_0,y_0) &= \int\left(\frac{d}{d\theta}m(\mathbf{x},\theta(F))\dot\theta(\mathbf{x}_0,y_0,F)f(\mathbf{x})\right)^2 d\mathbf{x} \\ &\quad + \big(y_0-m(\mathbf{x}_0,\theta(F))\big)^2 \\ &\quad - 2\big(y_0-m(\mathbf{x}_0,\theta(F))\big)\frac{d}{d\theta}m(\mathbf{x}_0,\theta(F)) \\ &\quad \times \dot\theta(\mathbf{x}_0,y_0,F)f(\mathbf{x}_0), \end{aligned} \tag{10}$$

where $\dot\theta(\mathbf{x}_0,y_0,F)$ is the influence function for the regression parameter estimator.

Clearly, $IF(\mathbf{x}_0,y_0)$ is not bounded in the y-direction. In contrast, $IF^*(\mathbf{x}_0,y_0)$ is bounded in the response space when the regression function is robustly estimated. The influence function analysis suggests that the robustified Zheng test will have more stable behavior than the original Zheng test when the response variable is contaminated. We note that neither test is bounded in the $\mathbf{X}$-direction, which can be seen clearly in the case of testing linearity where $\dot\theta(\mathbf{x}_0,y_0,F)$ is proportional to $\mathbf{x}_0$. Thus both tests will not be stable in the presence of bad leverage points.

4.3 Asymptotic Level Under a Shrinking Contaminated Neighborhood

Similar to Heritier and Ronchetti (1994), we consider the following shrinking point mass contamination:

$$F_{\epsilon,n}=\left(1-\frac{\epsilon}{\sqrt{nh^{l/2}}}\right)F+\frac{\epsilon}{\sqrt{nh^{l/2}}}\Delta_{(\mathbf{x}_0,y_0)},$$

where F is a distribution function in compliance with H_0. The rate at which the local contamination converges to 0 is the same as that of the local alternative sequences. This rate is slower than the parametric rate $n^{-1/2}$. This new feature is due to the intrinsic properties of smoothing. The von Mises functional expansion can be used to derive the asymptotic distribution under shrinking contamination.

Proposition 1. Assume the conditions of Theorem 1, under contamination $F_{\epsilon,n}$, if $h\to 0$, $nh^l\to\infty$, and $nh^{5l}\to 0$. Then

(a) $nh^{l/2}V_n\to \mathrm{N}(\epsilon^2 IF(\mathbf{x}_0,y_0),\tau^2)$ in distribution, where $\tau^2=2\sigma^4\int K^2(u)\,du\int f_{\mathbf{X}}^2(\mathbf{x})\,d\mathbf{x}$ and

(b) $nh^{l/2}V_n^*\to \mathrm{N}(\epsilon^2 IF^*(\mathbf{x}_0,y_0),\tau^{*2})$ in distribution, where $\tau^{*2}=\frac{1}{72}\int K^2(u)\,du\int f_{\mathbf{X}}^2(\mathbf{x})\,d\mathbf{x}$.

Let α_0 be the prespecified nominal level, and let $\alpha(F_{\epsilon,n})$ and $\alpha^*(F_{\epsilon,n})$ be the true levels of the Zheng test and the robustified Zheng test when the random sample is from the contaminated distribution $F_{\epsilon,n}$. Application of Taylor expansion leads to

$$\begin{aligned} &\lim_{n\to\infty}\alpha(F_{\epsilon,n}) \\ &\quad =\alpha_0+\epsilon^2\phi(\Phi^{-1}(1-\alpha))IF(\mathbf{x}_0,y_0)/\tau+o(\epsilon^2) \end{aligned} \tag{11}$$

and

$$\begin{aligned} &\lim_{n\to\infty}\alpha^*(F_{\epsilon,n}) \\ &\quad =\alpha_0+\epsilon^2\phi(\Phi^{-1}(1-\alpha))IF^*(\mathbf{x}_0,y_0)/\tau^*+o(\epsilon^2), \end{aligned} \tag{12}$$

where the $\Phi^{-1}(\cdot)$'s are the quantile function and the density function for the standard normal distribution. Therefore, the bias of the asymptotic level is bounded in the response space only if the second-order Hampel influence function of the test is bounded. This explains, from another angle, the unstable behavior of the original smoothing test and the relative robustness of the rank test. From the form of $IF^*(\mathbf{x}_0,y_0)$, we see that the rank transformation must be applied to residuals from a robust fit, such as an M-regression, for the robust test to have stable level behavior.

5. SIMULATIONS AND DATA EXAMPLE

5.1 Numerical Simulations

We consider testing nonlinearity of a regression function. The hypothesis of interest is

$$H_0: r(x)=a+bx \qquad \text{versus} \qquad H_1: r(x)\neq a+bx.$$

The results are based on 10,000 simulations. The Monte Carlo error of the simulated level is about $\sqrt{.05\times .95/10{,}000}\approx .002$. All data were generated using the statistical software R. The nominal level is taken to be .05. To calculate the Zheng test, the parameters are estimated using the least squares method, and the variance is estimated using Zheng's formula (3.9). To calculate the robust test, the residuals are obtained from a robust regression using the method-of-moments estimator of Yohai (1987), fitted by the "rlm" function from the MASS library in R. For both tests, the biquadratic kernel $K(u)=\frac{15}{16}(1-u^2)^2 I(|u|\le 1)$ is used. Our simulation experience suggests that the test is not sensitive to the choice of kernel function. (For other alternative choices of kernel function, see, e.g., Simonoff 1996, chap. 3.)

Example 1: Level of the Tests. The covariate X_i's are uniformly distributed on $(-2,2)$, and the response Y_i's are generated by $Y_i=1+2x_i+\epsilon_i$, $i=1,\dots,n$. Four simulation settings are considered: (a) the ϵ_i's are iid N(0, 1); (b) the ϵ_i's are iid from a log-normal distribution that is standardized to have mean 0 and variance 1; (c) the ϵ_i's are iid from N(0, 1), and 10% of the responses are randomly replaced by an outlying value 5; and (d) the ϵ_i's are iid from N(0, 1), and 10% of the responses are randomly replaced by observations from a nonlinear model $Y_i=5.5\cos(3\pi x)+\epsilon_i$. We denote the foregoing four settings S_1, S_2, S_3, and S_4 for convenience.

The observed size for the Zheng test and the robustified Zheng test are summarized in Table 1 for two different sample sizes, $n=60$ and 100, and four different values of the smoothing parameter h: .06, .09, .12, and .15.

Table 1. Empirical Level of Zheng's Test and the Robust Test

		Sample size n = 60				*Sample size n = 100*			
Test	*h*	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4
Zheng	.06	.049	.045	.065	.079	.051	.052	.067	.125
	.09	.046	.046	.059	.082	.048	.049	.064	.130
	.12	.046	.042	.056	.082	.042	.043	.057	.133
	.15	.044	.039	.051	.082	.040	.040	.056	.134
Robust	.06	.037	.045	.046	.052	.040	.044	.043	.057
	.09	.034	.043	.046	.049	.036	.042	.040	.062
	.12	.033	.041	.041	.048	.034	.039	.041	.062
	.15	.032	.040	.037	.045	.032	.038	.042	.060

For simulation settings S_1 and S_2, both tests are slightly conservative but are close to the specified nominal level. In general, nonparametric smoothing tests converge to their asymptotic normal distributions quite slowly (see, e.g., Härdle and Mammen 1993). The Zheng test does not require specification of the error distribution and thus has an asymptotically correct level under log-normal error. The loss of power of the Zheng test under nonnormality is shown in Example 2. When the response is contaminated, as in S_3 and S_4, the Zheng test becomes liberal (i.e., the observed level is as large as .13 for S_4), whereas the robustified Zheng test remains stable.

To give an idea of the computation complexity, the time taken to generate the entries in the column corresponding to simulation setting S_1 in Table 1 for both tests and two different sample sizes was about 11 minutes on a Pentium 1.60-GHz personal computer.

Example 2: Power of the Tests. In this example, we show the following: (a) the robustified Zheng test can be more powerful than the Zheng test in the presence of heavy-tailed errors, and (b) compared with classical robust tests, the robustified Zheng test can be more powerful for detecting various alternatives. More specifically, we compare this test with the rank-based drop-in-dispersion test (McKean and Schrader 1980; see also Hettmansperger and McKean 1998, sec. 3.6). In contrast to the robustified Zheng test, the drop-in-dispersion test can test only a subhypothesis; that is, an alternative model (in our simulations, it is taken to be a quadratic model) must be specified. The drop-in-dispersion test is calculated using the R software recently developed by Terpstra and McKean (2005).

We consider the following different simulation settings, which include different functional forms and different covariate and error distributions: (a) X_i is from uniform $(-2, 2)$ and the response is $Y_i = 1 + \theta X_i^2 + \epsilon_i$, where the ϵ_i's are iid from $N(0, 1)$; (b) X_i is from $N(0, 1)$ and the response is $Y_i = 1 + \cos(\theta\pi X_i) + \epsilon_i$, where the ϵ_i's are iid from a t distribution with 4 degrees of freedom that is standardized to have mean 0 and variance 1; (c) X_i is from $N(0, 1)$ and the response is $Y_i = \frac{10}{1+\theta\exp(-2X_i)} + \epsilon_i$, where the ϵ_i's are iid from a lognormal distribution that is standardized to have mean 0 and variance 1; and (d) X_i is from uniform $(0, 1)$ and the response is $Y_i = 1 - X_i + \theta\exp\{-200(X_i - .5)^2\} + \epsilon_i$, where the ϵ_i's are iid from a normal mixture distribution $.95N(0, 1) + .05N(10, 1)$ that is standardized to have mean 0 and variance 1.

The functional form in case (b) is used for modeling periodicity. The logistic function has a sigmoidal shape and is popular for modeling growth. The exponential form may arise from the solution of a differential equation. The simulated power curves for different values of θ and sample size $n = 100$ are displayed in Figure 1. The first simulation setting assumes the quadratic alternative and is designed to the advantage of the drop-in-dispersion test, which does exhibit the highest power in this case, although when $\theta > .6$, the power difference among the three tests becomes negligible. This simulation is also to the advantage of the Zheng test, because the random errors are normal. The Zheng test appears to be slightly more powerful than its robust alternative. For the cosine alternative, the robustified Zheng test has higher power than the other two tests. In fact, the power of the drop-in-dispersion test decreases dramatically, and the Zheng test loses power due to the heavier-tailed random errors. The same is observed for the third simulation setting, where the alternative is a logistic function. For the exponential alternative, the robust test exhibits the highest power, whereas the Zheng test has lower power than the drop-in-dispersion test.

To get a sense of the effect of sample size and bandwidth, we conduct a more in-depth analysis for the logistic alternative. We repeat the simulations for the three tests for $n = 60$ and display the simulated power curves in Figure 2(a). The patterns are similar as those for $n = 100$. Figures 2(b) and 2(c) give the simulated power curves for the robustified Zheng's test for different choices of h: .06, .09, .12, and .15 for $n = 60$ and $n = 100$. It appears that for $n = 100$, the power curves for the four different choices of h are very close to one another; for $n = 60$, smaller bandwidth corresponds to lower power. Figure 2(d) gives the empirical receiver operating curve (ROC) for three tests when $n = 100$, $h = .09$, and $\theta = .8$ for relevant nominal size range $[0, .2]$. The ROC curve compares power while adjusting for the difference in observed level (Lloyd 2005). In this example, for nominal level .05, the unadjusted power for the Zheng test, the robust Zheng test, and the drop-in-dispersion test are .942, .998, and .520, and the adjusted power is .930, .998, and .484. The difference appears to be negligible.

The problem of choosing a smoothing parameter h to optimize the power remains an open problem faced by all smoothing-based tests. It is widely recognized that the optimal choice of h for nonparametric curve estimation is generally not optimal for testing. A uniformly most powerful test does not exist due to the obstacle that there are infinitely many alternatives. King, Hart, and Wehrly (1991) suggested reporting the test results for different choices of h. Azzalini and Bowman (1991) proposed plotting the p value against h and called this technique "smoothing trace." Horowitz and Spokoiny (2001) circumvented the problem of choosing a single h by using the maximum of a studentized smoothing conditional moment test over a sequence of smoothing parameters and proved that this leads to a test with a certain optimality property. Dette and Hetzler (2004) treated smoothing conditional moments test indexed by bandwidths as empirical processes. Generalization of either approach to the current context of robust testing requires substantial work and is left for future research.

5.2 An Agricultural Data Example

The relationship between the yield of a crop and the density of planting is of great interest in agriculture. Here we analyze a dataset from Ratkowsky (1983) comprising of 42 observations on the yield (g/plant), on the log scale, of onions in Virginia, South Australia. The covariate is the density of planting (plants/m^2). A scatterplot of this dataset is given in Figure 3(a).

We examine the possible nonlinearity using the tests discussed in this paper. The "significance trace" of the Zheng test and the robust Zheng test are illustrated in Figure 3(b). We note that except for very small smoothing parameters (which are not recommended), both give very small p values (close to 0). The drop-in-dispersion test also gives a very small p value (.007). All three tests indicate that the linearity assumption should be rejected.

To investigate the influence of a single outlier in the response space, we artificially change $y(35)$, the response of the 35th observation pair $(101.81, 4.11)$, from 4.11 to 3. The "significance

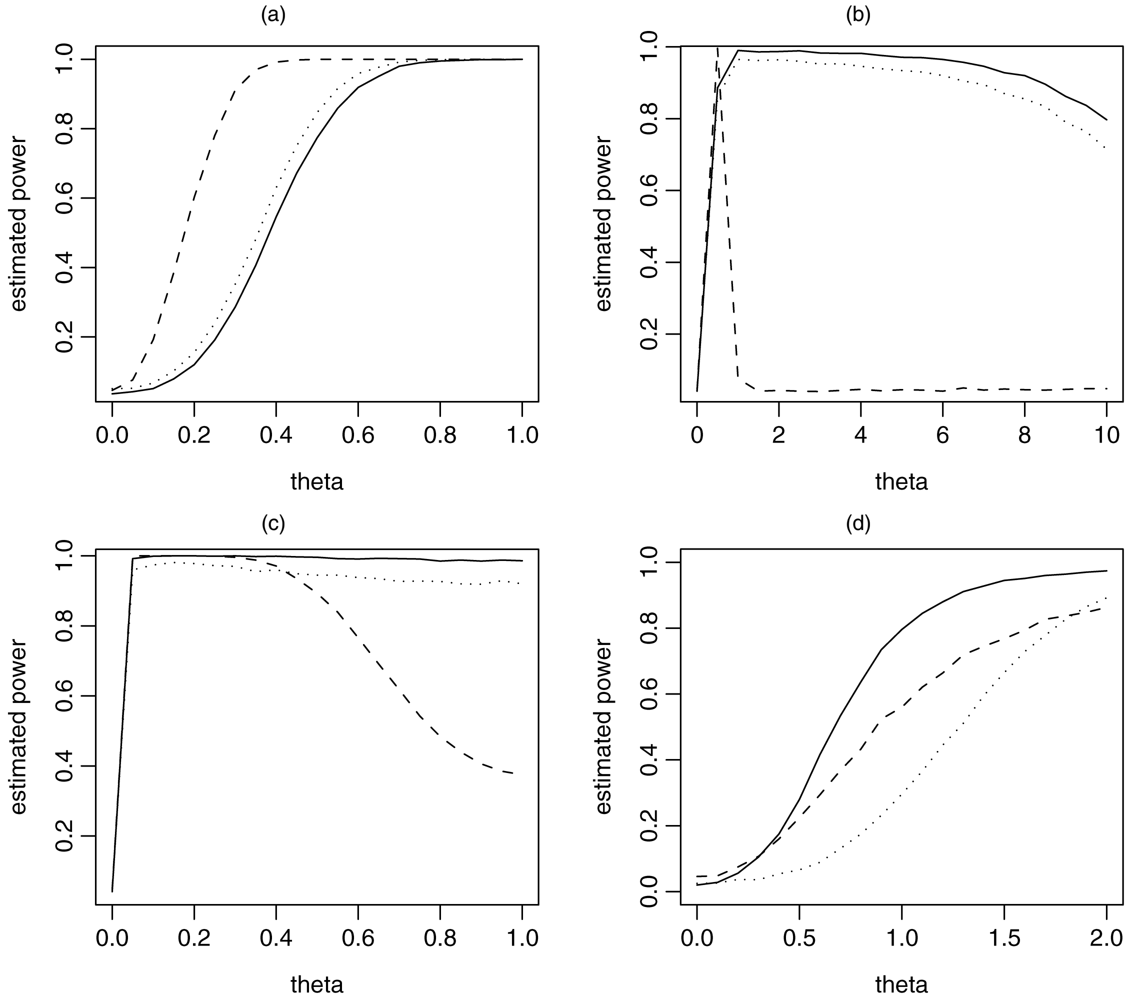

Figure 1. Simulated Power Curves for (a) Quadratic Alternative, (b) Cosine Alternative, (c) Logistic Alternative, and (d) Exponential Alternative ($n = 100$, $h = .09$; ——, robust test; · · · ·, Zheng's test; – –, drop-in-dispersion test).

trace" of the Zheng test and the robust Zheng test corresponding to this change are given in Figure 3(b). We observe that the p value of the Zheng test quickly becomes much greater than the specified nominal level, whereas the robust alternative is only slightly influenced. The effect of the outlier on the Zheng test is quite dramatic considering that the change from 4.11 to 3 represents a relatively small contamination, because the range of the response variable for this dataset is $(3, 6)$. The drop-in-dispersion test is also only slightly influenced, with a p value of around .004.

We next perform a power breakdown analysis by adding c ($c = 4, 6, 8, 10$) copies of $(101.81, 5)$ to the original data and carry out the same tests. The "significance trace" of the Zheng test and the robust Zheng test are given in Figure 4(a). It is clear that the Zheng test is seriously affected even for $c = 4$, whereas the power of the robust Zheng test remains robust when $c = 8$ (i.e., the contamination is 16%) and becomes affected when $c = 10$. Our simulations also indicate that the drop-in-dispersion test begins to fail to reject the linearity null hypothesis at values of $c = 4$ and above. The p values for the drop-in-dispersion test for $c = 2$ and 3 are .0375 and .075.

Ratkowsky suggested fitting the nonlinear yield-density model $Y_i = -\log(\alpha + \beta X_i) + \epsilon_i$, which was referred to as the reciprocal model by Mead (1970). Both the Zheng test and the robustified Zheng test give p values much higher than .05 and thus do not reject this choice. We study level breakdown by artificially introducing c ($c = 2, 4, 6, 8$) copies of $(101.81, 5)$ into the data. The results of the Zheng test and the robust Zheng test are given in Figure 4(b). The level breakdown occurs for the Zheng test when $c = 4$ (i.e., the contamination is about 9%), and occurs much later for the robust Zheng test, when $c = 8$ (i.e., the contamination is 16%). For this last analysis, the nonlinear regression for the Zheng test is fitted using the function nls in R, and the nonlinear robust regression for the robust test is fitted using the function nlrob from the library robustbase in R. The drop-in-dispersion test in the software developed by Terpstra and McKean (2005) is only for linear models and thus is not applied here.

6. GENERALIZATIONS AND DISCUSSION

6.1 Other Smoothing Conditional Moment Tests

The smoothing-based conditional moment tests include many existing methods, including those listed in Section 1. Härdle, Müller, Sperlich, and Werwatz (2004, p. 128) pointed out that many statistics can be considered estimates of some com-

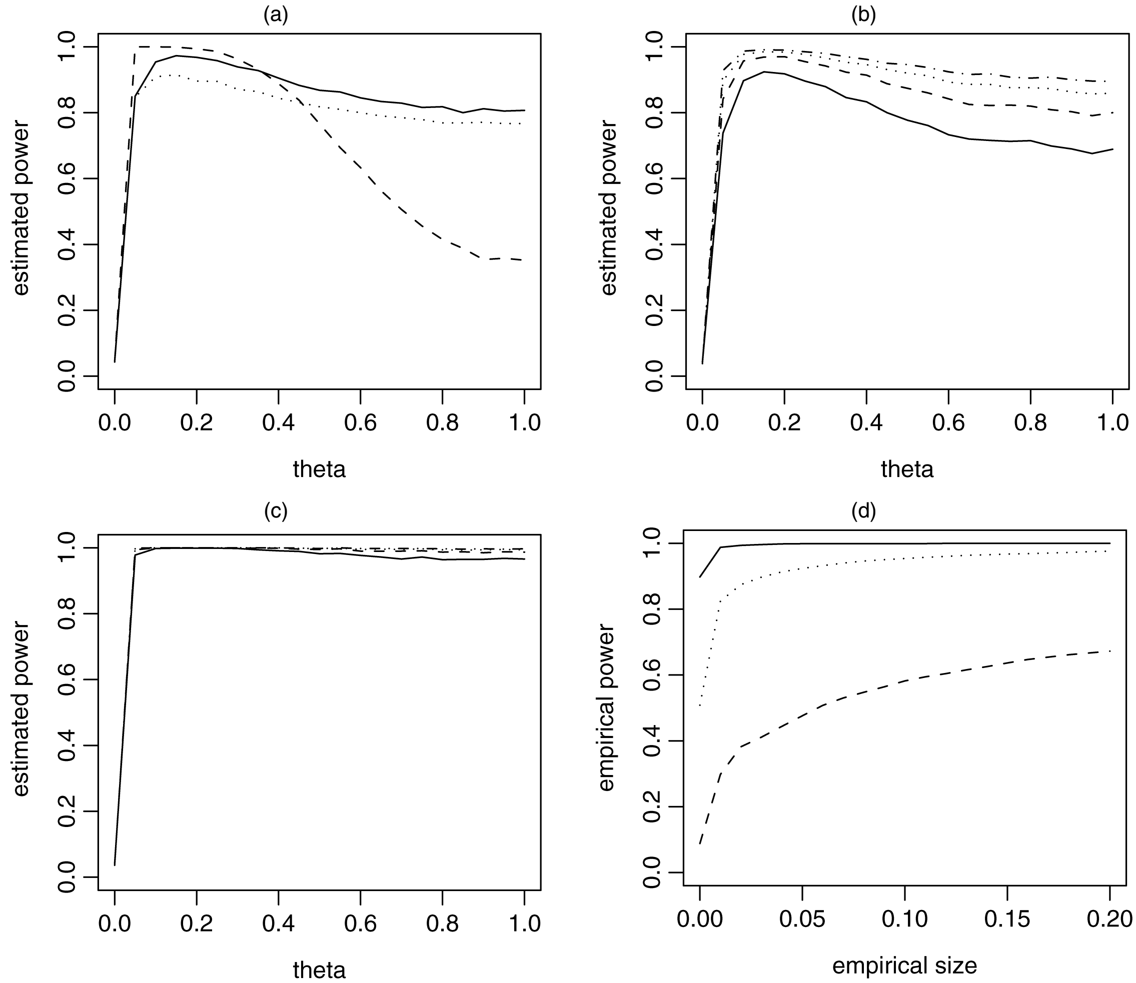

Figure 2. Simulated Power Curves of the Logistic Alternative. (a) Simulated power of the three tests for $n = 60$ and $h = .09$ ($\cdots$, Zheng test; ——, robust Zheng test; – –, drop-in-dispersion test). (b) Simulated power of the robust Zheng test for $n = 60$ and $h = .06$ (——), .09 (– –), .12 ($\cdots$), .15 (· – ·). (c) Simulated power of the robust Zheng test for $n = 100$ and $h = .06$ (——), .09 (– –), .12 ($\cdots$), .15 (· – ·). (d) Empirical ROC curves for the three tests for $n = 100$, $h = .09$, and $\theta = .8$, restricted to relevant nominal size range [0, .2] ($\cdots$, Zheng test; ——, robust Zheng test; – –, drop-in-dispersion test).

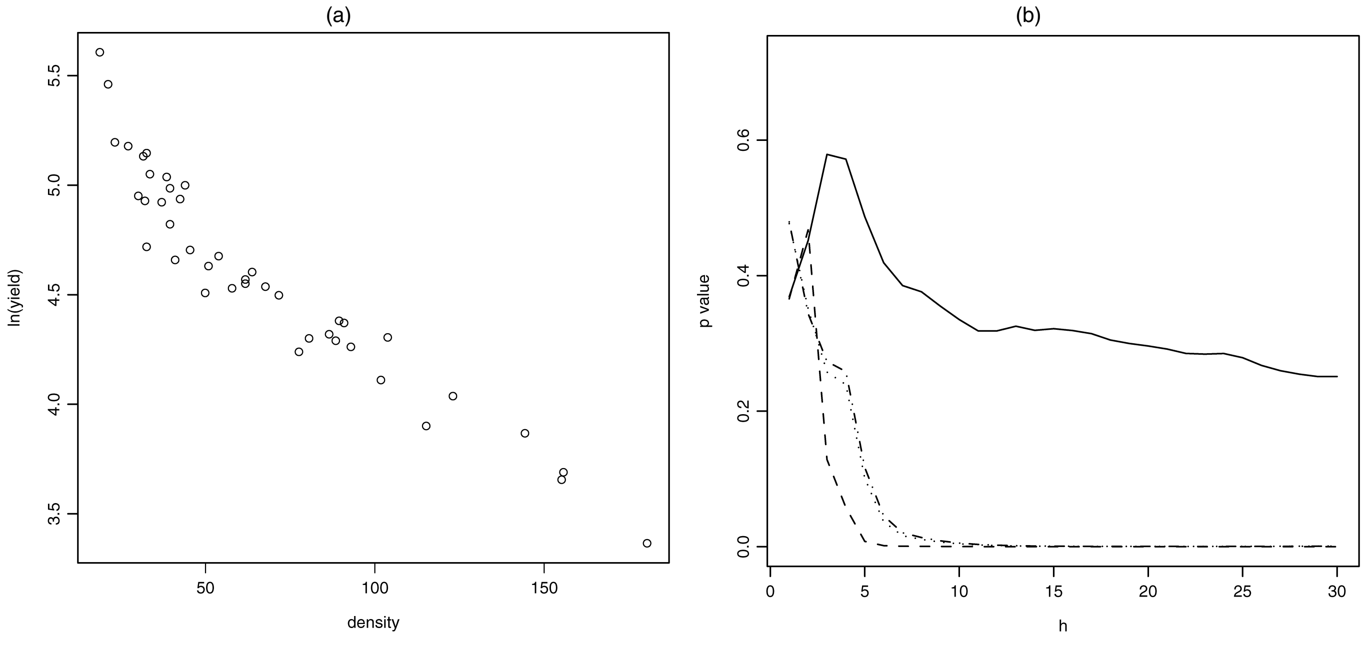

Figure 3. Onion Data. (a) Scatterplot. (b) Testing linearity using the Zheng test and the robust Zheng test with and without outliers [– – Zheng: original data; —— Zheng: $y(35) = 3$; $\cdots$ Robust: original data; · – · Robust; $y(35) = 3$].

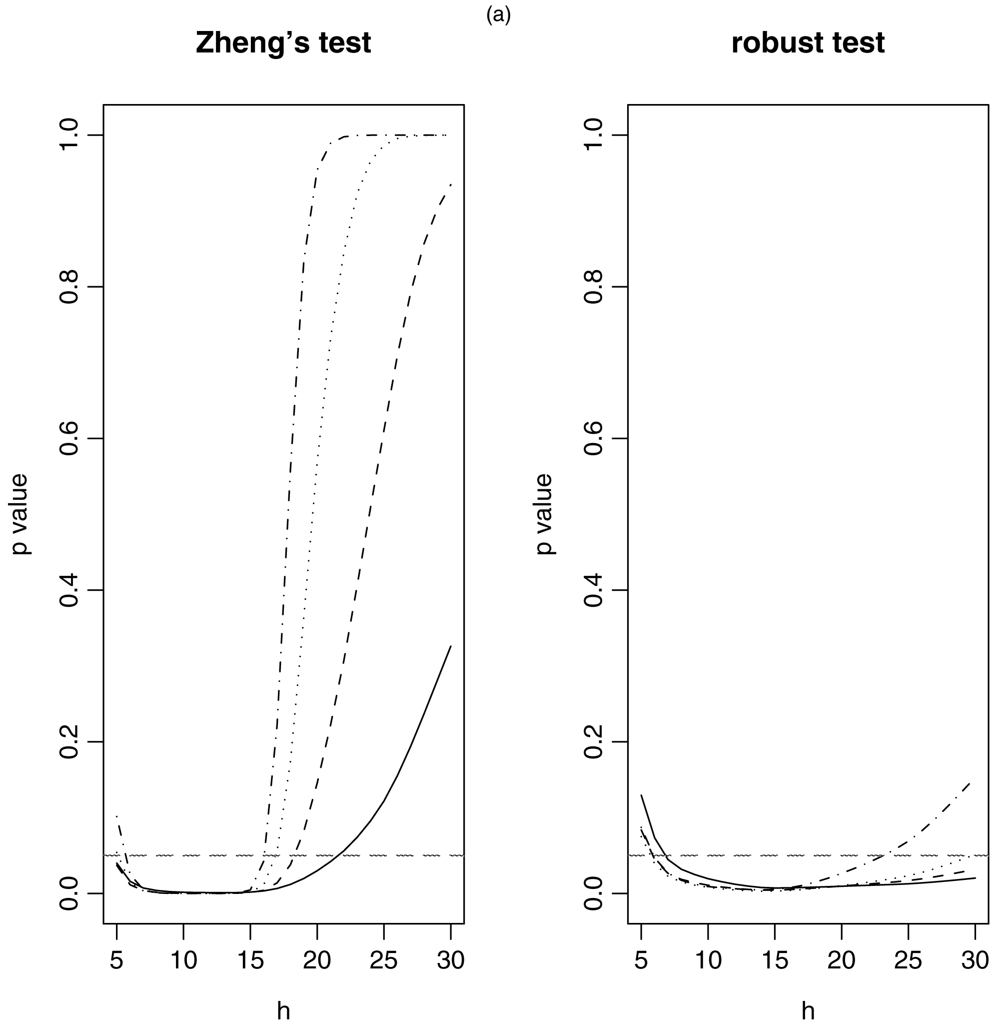

Figure 4. Onion Data: Breakdown Analysis. (a) Power breakdown analysis: Null model is a linear model. (——, c = 4; – –, c = 6; · · · ·, c = 8; · – ·, c = 10.) The horizontal dashed line represents p = .05.

mon moment conditions. Zhang and Dette (2004) made similar observations.

We have focused on the robustified Zheng test in this article due to its technical tractability and easy computation. The approach is readily applicable to many other smoothing-based conditional moments tests. For instance, the test statistic of Härdle and Mammen (1993) is $\int [\widehat{r}(\mathbf{x}) - \widehat{m}(\mathbf{x}, \widehat{\theta})]^2 w(\mathbf{x})\, d\mathbf{x}$, where $w(\mathbf{x})$ is some nonnegative weight function and $\widehat{r}(\mathbf{x}) - \widehat{m}(\mathbf{x}, \widehat{\theta})$ is the smoothed version of the residual under H_0. We expect that replacing the residuals by their asymptotic rank transformations will lead to robustification.

6.2 A Modification of the Robust Test

To obtain a test with influence function bounded in the whole $(\mathbf{X}, Y)$ space, we consider the following simple modification of V_n^* by introducing a weight function in the $\mathbf{X}$ space,

$$V_n^* = \frac{1}{n(n-1)} \sum_{j_1=1}^{n} \sum_{j_2 \neq j_1} \frac{1}{h^l} K\left(\frac{\mathbf{X}_{j_1} - \mathbf{X}_{j_2}}{h}\right) \widehat{\epsilon}_{j_1}^* \widehat{\epsilon}_{j_2}^* \pi(\mathbf{X}_{j_1}),$$

where $\pi(\cdot)$ is a positive weight function. The corresponding Hampel first-order influence function is zero, and the second-order influence function becomes

$$\begin{aligned} &IF^*(\mathbf{x}_0, y_0) \\ &\quad = \int \left[\int \dot{G}^*_{\Delta_{(\mathbf{x}_0,y_0)}} \{y - m(\mathbf{x}, \theta(F))\} f(\mathbf{x}, y) \pi(\mathbf{x})\, dy \right]^2 d\mathbf{x} \\ &\qquad + 2\{H(y_0 - m(\mathbf{x}_0, \theta(F))) - 1/2\} \\ &\qquad \times \int \dot{G}^*_{\Delta_{(\mathbf{x}_0,y_0)}} \{y - m(\mathbf{x}_0, \theta(F))\} f(\mathbf{x}_0, y)\, dy\, \pi(\mathbf{x}_0) \\ &\qquad + \{H(y_0 - m(\mathbf{x}_0, \theta(F))) - 1/2\}^2 \pi(\mathbf{x}_0). \qquad (13) \end{aligned}$$

It is clear that an appropriate choice of $\pi(\cdot)$ would lead to an influence function bounded in the whole $(\mathbf{X}, Y)$ space. For example, we may take $\pi(\mathbf{x}_0) = \min(1, \xi^{-1}(\mathbf{x}_0))$, where $\xi(\mathbf{x}_0) = \int \dot{G}^*_{\Delta_{(\mathbf{x}_0,y_0)}} \{y - m(\mathbf{x}_0, \theta(F))\} f(\mathbf{x}_0, y)\, dy$. Under regularity conditions as before, the modified test still has an asymptotic normal distribution.

6.3 Summary and Discussion

We have proposed a robust smoothing-based conditional moment test for diagnosing the functional form of a regression function. This test is consistent against any alternative in an

(b)

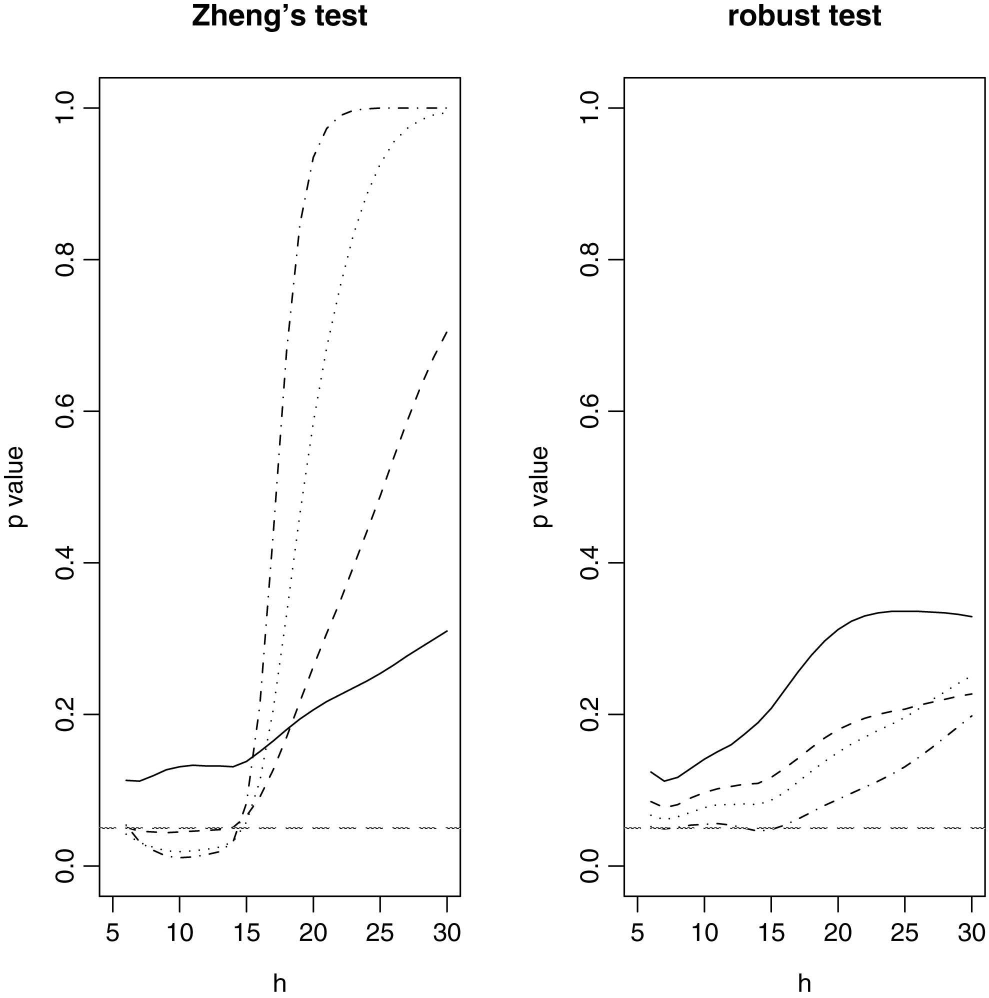

Figure 4 (Continued). Onion Data: Breakdown Analysis. (b) Level breakdown analysis: Null model is the reciprocal model. (——, c = 2; − −, c = 4; · · · ·, c = 6; · − ·, c = 8.) The horizontal dashed line represents p = .05.

infinite-dimensional functional space. The new approach provides a bridge connecting existing robust tests and the more up-to-date smoothing-based tests.

The smoothing test may have limited application if the data contain a large portion of sparse data or the sample size is very small. An alternative idea for developing a test with omnibus alternative and bounded influence is to use a robust smoother (see Cleveland 1979; Härdle and Gasser 1984; Tsybakov 1986; Fan, Hu, and Truong 1994; Welsh 1996). For instance, we may define a test statistic as a distance measure between a parametric M regression and an M smoother. However, because an M smoother does not have a closed-form expression, studying the asymptotic theory is much more challenging. A possible future research direction is to investigate whether any bounds on the breakdown points of the level of the robust test can be derived as was done by He, Simpson, and Portnoy (1990) and Markatou and He (1994).

APPENDIX: PROOFS

Here we give an outline of the proof of Theorem 1 and formula (8). Proofs of Theorem 2, Theorem 3, and Proposition 1 are available in a separate technical report.

Proof of Theorem 1

It is sufficient to show that $nh^{l/2}(V_n^* - V_n^{**}) \xrightarrow{p} 0$. We have

$$
\begin{aligned}
V_n^* - V_n^{**} &= \frac{1}{n^3(n-1)h^l} \sum_{j_1=1}^{n} \sum_{j_2 \neq j_1}^{n} \sum_{j_3=1}^{n} \sum_{j_4=1}^{n} K\left(\frac{\mathbf{X}_{j_1} - \mathbf{X}_{j_2}}{h}\right) \\
&\quad \times \left[I(\widehat{\epsilon}_{j_3} \le \widehat{\epsilon}_{j_1}) - I(\epsilon_{j_3} \le \epsilon_{j_1})\right] \\
&\quad \times \left[I(\widehat{\epsilon}_{j_4} \le \widehat{\epsilon}_{j_2}) - I(\epsilon_{j_4} \le \epsilon_{j_2})\right] \\
&\quad + \frac{2}{n^3(n-1)h^l} \sum_{j_1=1}^{n} \sum_{j_2 \neq j_1}^{n} \sum_{j_3=1}^{n} \sum_{j_4=1}^{n} K\left(\frac{\mathbf{X}_{j_1} - \mathbf{X}_{j_2}}{h}\right) \\
&\quad \times \left[I(\widehat{\epsilon}_{j_3} \le \widehat{\epsilon}_{j_1}) - I(\epsilon_{j_3} \le \epsilon_{j_1})\right]\left[I(\epsilon_{j_4} \le \epsilon_{j_2}) - \frac{n+1}{2n}\right] \\
&\quad + \frac{1}{n^3(n-1)h^l} \sum_{j_1=1}^{n} \sum_{j_2 \neq j_1}^{n} \sum_{j_3=1}^{n} \sum_{j_4=1}^{n} K\left(\frac{\mathbf{X}_{j_1} - \mathbf{X}_{j_2}}{h}\right) \\
&\quad \times \left[\left(I(\epsilon_{j_3} \le \epsilon_{j_1}) - \frac{n+1}{2n}\right)\left(I(\epsilon_{j_4} \le \epsilon_{j_2}) - \frac{n+1}{2n}\right)\right. \\
&\quad \left. - \left(H(\epsilon_{j_1}) - 1/2\right)\left(H(\epsilon_{j_2}) - 1/2\right)\right] \\
&= D_1 + D_2 + D_3,
\end{aligned}
$$

where the definition of D_i, $i=1,2,3$, should be clear from the context. To show that $nh^{l/2}D_1=o_p(1)$, we first verify a useful probability bound,

$$\max_{1\le j_1\le n}\sum_{\substack{j_3=1\\ j_3\ne j_1}}^{n}\bigl|I(\widehat{\epsilon}_{j_3}\le\widehat{\epsilon}_{j_1})-I(\epsilon_{j_3}\le\epsilon_{j_1})\bigr|=O_p(n^{1/2}),\qquad\text{(A.1)}$$

where $\widehat{\epsilon}_i=\epsilon_i+(f(\mathbf{X}_i,\theta_0)-f(\mathbf{X}_i,\widehat{\theta}_n))$. Let $L(\theta^*)$ be the sum on the left side of (A.1) with $\widehat{\theta}_n$ replaced by θ^*. Denote $B=\{\theta^*:n^{1/2}|\theta^*-\theta_0|\le\delta$ for $\delta=O(1)\}$, $Z_{j_1j_3}=\epsilon_{j_3}-\epsilon_{j_1}$, and $t(\mathbf{X}_{j_1},\mathbf{X}_{j_3},\theta,\theta^*)=(m(\mathbf{X}_{j_1},\theta_0)-m(\mathbf{X}_{j_1},\theta^*))-(m(\mathbf{X}_{j_3},\theta)-m(\mathbf{X}_{j_3},\theta^*))$; then

$$\begin{aligned}\sup_{\theta^*\in B}|L(\theta^*)|&=\sup_{\theta^*\in B}\sum_{\substack{j_3=1\\ j_3\ne j_1}}^{n}\bigl|I(Z_{j_1j_3}\le t(\mathbf{X}_{j_1},\mathbf{X}_{j_3},\theta,\theta^*))-I(Z_{j_1j_3}\le 0)\bigr|\\&\le\sup_{\theta^*\in B}\sum_{\substack{j_3=1\\ j_3\ne j_1}}^{n}I(|Z_{j_1j_3}|\le|t(\mathbf{X}_{j_1},\mathbf{X}_{j_3},\theta,\theta^*)|)\\&\le\sum_{\substack{j_3=1\\ j_3\ne j_1}}^{n}I(|Z_{j_1j_3}|\le Cn^{-1/2}).\end{aligned}$$

Conditional on ϵ_{j_1}, $I(|Z_{j_1j_3}|\le Cn^{-1/2})$, $j_3\ne j_1$, are iid Bernoulli random variables with success probability of order $O(n^{-1/2})$, and C is used in this proof to denote a generic positive constant, which may change from line to line. An application of Bernstein's inequality leads to $P[\sum_{j_3=1,j_3\ne j_1}^{n}I(|Z_{j_1j_3}|\le Cn^{-1/2})\ge Cn^{1/2}|\epsilon_{j_1}]\le\exp(-Cn^{1/2})$. Unconditionally, we also have that

$$P\Biggl[\sum_{j_3\ne j_1}^{n}I(|Z_{j_1j_3}|\le Cn^{-1/2})\ge Cn^{1/2}\Biggr]\le\exp(-Cn^{1/2});$$

therefore,

$$P\Biggl[\max_{1\le j_1\le n}\sum_{\substack{j_3=1\\ j_3\ne j_1}}^{n}I(|Z_{j_1j_3}|\le Cn^{-1/2})\ge Cn^{1/2}\Biggr]\le n\exp(-Cn^{1/2}).$$

Thus $\max_{1\le j_1\le n}\sum_{j_3\ne j_1}^{n}I(|Z_{j_1j_3}|\le Cn^{-1/2})=O_p(n^{1/2})$, and (A.1) holds. Now we have

$$\begin{aligned}nh^{l/2}|D_1|&\le\frac{1}{n^2(n-1)h^{l/2}}\max_{1\le j_1\le n}\sum_{\substack{j_3=1\\ j_3\ne j_1}}^{n}\bigl|I(\widehat{\epsilon}_{j_3}\le\widehat{\epsilon}_{j_1})-I(\epsilon_{j_3}\le\epsilon_{j_1})\bigr|\\&\quad\times\max_{1\le j_2\le n}\sum_{\substack{j_4=1\\ j_2\ne j_2}}^{n}\bigl|I(\widehat{\epsilon}_{j_4}\le\widehat{\epsilon}_{j_2})-I(\epsilon_{j_4}\le\epsilon_{j_2})\bigr|\\&\quad\times\sum_{j_1=1}^{n}\sum_{j_2\ne j_1}^{n}K\Bigl(\frac{\mathbf{X}_{j_1}-\mathbf{X}_{j_2}}{h}\Bigr)\\&=O_p(n^{-2}h^{-l/2})\sum_{j_1=1}^{n}\sum_{j_2\ne j_1}^{n}K\Bigl(\frac{\mathbf{X}_{j_1}-\mathbf{X}_{j_2}}{h}\Bigr).\end{aligned}$$

Let $U_n=\frac{1}{n(n-1)}\sum_{j_1=1}^{n}\sum_{j_2\ne j_1}^{n}\frac{1}{h^{l/2}}K(\frac{\mathbf{X}_{j_1}-\mathbf{X}_{j_2}}{h})$; then this is a U-statistic with a sample-dependent kernel. It is straightforward to check whether $E(\frac{1}{h^l}K^2(\frac{\mathbf{X}_{j_1}-\mathbf{X}_{j_2}}{h}))=O(1)$. The condition of lemma 3.1 of Zheng (1996) is satisfied, and we have $U_n=\overline{r}_n+o_p(1)$, where $\overline{r}_n=E(\frac{1}{h^{l/2}}K(\frac{\mathbf{X}_{j_1}-\mathbf{X}_{j_2}}{h}))=o(1)$. Thus we have shown that $nh^{l/2}|D_1|\le o_p(1)$. Similarly, we can show that $nh^{l/2}|D_i|=o_p(1)$ for $i=2,3$.

Verification of (8)

Write $F_{1t}=F+tG_1$ and $F_{2t}=F+tG_2$, and let f_{1t} and g_1 be the density functions corresponding to F_{1t} and G_1. We have

$$\begin{aligned}A(t)&=\iint[G_t^*(y-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times\Bigl(\int[G_t^*(y_1-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times f_{1t}(\mathbf{x},y_1)\,dy_1\Bigr)dF_{2t}(\mathbf{x},y)\\&=\iint[G_t^*(y-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times\Bigl(\int[G_t^*(y_1-m(\mathbf{x},\theta(F_{2t})))-1/2]f(\mathbf{x},y_1)\,dy_1\Bigr)dF(\mathbf{x},y)\\&\quad+t\iint[G_t^*(y-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times\Bigl(\int[G_t^*(y_1-m(\mathbf{x},\theta(F_{2t})))-1/2]g_1(\mathbf{x},y_1)\,dy_1\Bigr)dF(\mathbf{x},y)\\&\quad+t\iint[G_t^*(y-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times\Bigl(\int[G_t^*(y_1-m(\mathbf{x},\theta(F_{2t})))-1/2]f(\mathbf{x},y_1)\,dy_1\Bigr)dG_2(\mathbf{x},y)\\&\quad+t^2\iint[G_t^*(y-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times\Bigl(\int[G_t^*(y_1-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times g_1(\mathbf{x},y_1)\,dy_1\Bigr)dG_2(\mathbf{x},y)\\&=A_1(t)+A_2(t)+A_3(t)+A_4(t),\end{aligned}$$

where the definition of $A_i(t)$, $i=1,\dots,4$, should be clear from the context. We have

$$\begin{aligned}\frac{dA_1(t)}{dt}&=\iint\frac{d}{dt}G_t^*(y-m(\mathbf{x},\theta(F_{2t})))\\&\quad\times\int[G_t^*(y_1-m(\mathbf{x},\theta(F_{2t})))-1/2]f(\mathbf{x},y_1)\,dy_1\,dF(\mathbf{x},y)\\&\quad+\iint[G_t^*(y-m(\mathbf{x},\theta(F_{2t})))-1/2]\\&\quad\times\int\frac{d}{dt}G_t^*(y-m(\mathbf{x},\theta(F_{2t})))f(\mathbf{x},y_1)\,dy_1\,dF(\mathbf{x},y).\end{aligned}$$

Because $\int[G^*(y-m(\mathbf{x},\theta(F)))-1/2]\,dF(y|\mathbf{x})=0$, we have that $\frac{dA_1(t)}{dt}|_{t=0}=0$. Thus

$$\begin{aligned}\frac{d^2A_1(t)}{dt^2}\bigg|_{t=0}&=2\iint\dot{G}^*(y-m(\mathbf{x},\theta(F)))\\&\quad\times\int\dot{G}^*(y_1-m(\mathbf{x},\theta(F)))f(\mathbf{x},y_1)\,dy_1\,dF(\mathbf{x},y).\end{aligned}$$

In the same manner, $\frac{dA_i(t)}{dt}|_{t=0}=0$ for $i=2,3,4$ and $\frac{1}{2}\frac{d^2A_2(t)}{dt^2}|_{t=0}$ for $i=2,3,4$ are equal to the other three terms in (8).

[Received March 2005. Revised July 2006.]

REFERENCES

Agostinelli, C., and Markatou, M. (2001), "Robust Tests Based on the Weighted Likelihood Methodology," *Statistica Sinica*, 11, 499–514.

Azzalini, A., and Bowman, A. W. (1991), "Nonparametric Methods for Repeated Measurements," in *Nonparametric Functional Estimation and Related Topics*, ed. G. Roussas, Dordrecht: Kluwer Academic Publishers, pp. 377–387.

Bickel, P. J. (1978), "Using Residuals Robustly I: Tests for Heteroscedasticity, Nonlinearity," *The Annals of Statistics*, 6, 266–291.

Cantoni, E., and Ronchetti, E. (2001), "Bounded Influence for Generalized Linear Models," *Journal of the American Statistical Association*, 96, 1022–1030.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Dette, H., and Hetzler, B. (2004), "Specification Tests Indexed by Bandwidths," technical report, Ruhr-Universität Bochum, Fakultät für Mathematik.

Fan, J. Q., Hu, T.-C., and Truong, Y. K. (1994), "Robust Nonparametric Function Estimation," *Scandinavian Journal of Statistics*, 21, 433–446.

Fan, J. Q., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and the Wilks Phenomenon," *The Annals of Statistics*, 29, 153–193.

Fernholz, L. T. (1983), *von Mises Calculus for Statistical Functionals*, New York: Springer-Verlag.

González-Manteiga, W., and Cao, C. (1993), "Testing the Hypothesis of a General Linear Model Using Nonparametric Regression Estimation," *Test*, 2, 161–188.

Hall, P. (1984), "Central Limit Theorem for Integrated Squared Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis*, 14, 1–16.

Hampel, F. R. (1974), "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley.

Härdle, W. (1992), *Applied Nonparametric Regression*, Cambridge, U.K.: Cambridge University Press.

Härdle, W., and Gasser, T. (1984), "Robust Non-Parametric Function Fitting," *Journal of the Royal Statistical Society*, Ser. B, 46, 42–51.

Härdle, W., and Mammen, E. (1993), "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004), *Nonparametric and Semiparametric Models*, Berlin: Springer-Verlag.

Hart, J. (1997), *Nonparametric Smoothing and Lack-of-Fit Test*, New York: Springer-Verlag.

He, X., Simpson, D. G., and Portnoy, S. L. (1990), "Breakdown Robustness of Tests," *Journal of the American Statistical Association*, 85, 446–452.

Heritier, S., and Ronchetti, E. (1994), "Robust Bounded-Influence Tests in General Parametric Models," *Journal of the American Statistical Association*, 89, 897–904.

Hettmansperger, T. P., and McKean, J. W. (1998), *Robust Nonparametric Statistical Methods*, London: Arnold.

Horowitz, J. L., and Spokoiny, V. G. (2001), "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599–631.

King, E. C., Hart, J. D., and Wehrly, T. E. (1991), "Testing the Equality of Regression Curves Using Linear Smoothers," *Statistics and Probability Letters*, 12, 239–247.

Li, Q., and Wang, S. (1998), "A Simple Consistent Bootstrap Test for a Parametric Regression Function," *Journal of Econometrics*, 87, 145–165.

Liu, R. C., Markatou, M., and Tsai, C. L. (2005), "Robust Estimation and Testing in Nonlinear Regression Models," *International Journal of Pure and Applied Mathematics*, 21, 525–552.

Lloyd, C. J. (2005), "On Comparing the Accuracy of Competing Tests of the Same Hypothesis From Simulation Data," *Journal of Statistical Planning and Inference*, 128, 497–508.

Markatou, M., and He, X. (1994), "Bounded-Influence and High-Breakdown-Point Testing Procedures in Linear Models," *Journal of the American Statistical Association*, 89, 543–549.

Markatou, M., and Hettmansperger, T. P. (1990), "Robust Bounded-Influence Tests in Linear Models," *Journal of the American Statistical Association*, 85, 187–190.

Markatou, M., and Manos, G. (1996), "Robust Tests in Nonlinear Regression Models," *Journal of Statistical Planning and Inference*, 55, 205–217.

McKean, J. W., and Schrader, R. M. (1980), "The Geometry of Robust Procedures in Linear Models," *Journal of the Royal Statistical Society*, Ser. B, 42, 366–371.

Mead, R. (1970), "Plant Density and Crop Yield," *Applied Statistics*, 19, 64–81.

Newey, W. K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.

Ratkowsky, D. A. (1983), *Nonlinear Regression Modelling: A Unified Practical Approach*, New York: Marcel Dekker.

Schrader, R. M., and Hettmansperger, T. P. (1980), "Robust Analysis of Variance Based Upon a Likelihood Ratio Criterion," *Biometrika*, 67, 93–101.

Sen, P. K. (1982), "On M-Tests in Linear Models," *Biometrika*, 69, 245–248.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, New York: Springer-Verlag.

Tauchen, G. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.

Terpstra, J., and McKean, J. (2005), "Rank-Based Analyses of Linear Models Using R," *Journal of Statistical Software*, 14, 1–26.

Tsybakov, A. B. (1986), "Robust Reconstruction of Functions by the Local-Approximation Method," *Problems of Information Transmission*, 22, 133–146.

von Mises, R. (1947), "On the Asymptotic Distribution of Differentiable Statistical Functions," *Annals of the Institute of Statistical Mathematics*, 18, 309–348.

Welsh, A. H. (1996), "Robust Estimation of Smooth Regression and Spread Functions and Their Derivatives," *Statistica Sinica*, 6, 347–366.

Yohai, V. J. (1987), "High-Breakdown-Point and High-Efficiency Robust Estimates for Regression," *The Annals of Statistics*, 15, 642–656.

Zhang, C. M., and Dette, H. (2004), "A Power Comparison Between Nonparametric Regression Tests," *Statistics and Probability Letters*, 66, 289–301.

Zheng, J. X. (1996), "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.